

A gamma-mixture class of distributions with Bayesian application

Janet van Niekerk^{1,*}, Andriëtte Bekker¹, Mohammad Arashi²

Abstract

In this paper a subjective Bayesian approach is followed to derive estimators for the parameters of the normal model by assuming a gamma-mixture class of prior distributions, that includes the gamma and the noncentral gamma as special cases. An innovative approach is proposed to find the analytical expression of the posterior density function when a complicated prior structure is ensued. The simulation studies and a real dataset illustrates the modeling advantages of this proposed prior and support some of the findings.

Keywords: Bayesian inference, hypergeometric gamma, mixture of gamma, normal-gamma, normal-inverse gamma, variance

AMS 2000 Classification codes: 62F15, 62H12, 62H10, 60E05

1. Introduction

Bayesian analysis of the normal model has been discussed by many authors each following a different approach. Objective and conjugate subjective priors were considered by [1] amongst others. [2] adopted the non-conjugate subjective normal-gamma prior. This normal-gamma prior has been showed to perform well in the multiple linear regression model by [3] when compared with the

*Corresponding author

Email address: janet.vanniekerk@up.ac.za (Janet van Niekerk)

¹Department of Statistics, Faculty of Natural and Agricultural Sciences, University of Pretoria, Pretoria, South Africa

²Department of Statistics, School of Mathematical Sciences, Shahrood University of Technology, Shahrood, Iran. Department of Statistics, Faculty of Natural and Agricultural Sciences, University of Pretoria, Pretoria, South Africa

Bayesian lasso estimator. As indicated by [4] and [5] that non-informative priors can have strong and undesirable implications for inference. Ignoring important information by assuming priors that cannot elicit that information will lead to inaccurate analysis and inference, hence it is important to assume a parameter-rich model that will enable one to model appropriate prior beliefs about the parameter of interest. Computational limitations within the Bayesian framework also no longer present a problem to analysing any data set with complex prior assumptions; a generalised prior can therefore be handled with ease in Bayesian inference. As with this prior it is desirable, as mention by Sebastian [6], to consider priors that exhibit flexibility regarding the shape of their tails.

The genesis for this paper followed from the performance study done by [7] to evaluate the subjective normal-gamma and normal-inverse gamma distributions as possible priors for the parameters of the normal model. It was demonstrated by comparing different measures for a simulated as well as a real dataset that the non-conjugate mathematically intractable normal-gamma is a suitable competitor for the conjugate normal-inverse gamma prior. In this paper, a flexible class that is the univariate counterpart emanating from the generalized matrix variate Wishart distribution of [8], is proposed as the prior for the variance of the normal model. An efficient computational approach is outlined in this paper and demonstrates the relative computational ease in assuming complex priors.

In Section 2, this gamma class is proposed, and a Bayesian analysis of the normal model is performed, assuming this class as the prior for the variance (see Section 3), by identifying complex integrals as expected values of functions of the parameters, simpler analytical expressions are obtained. Section 4 consists of some illustrations regarding this new prior structure, and the performance is measured by the bias, mean squared error (MSE) and coverage probabilities including two other subjective priors and the MLE's. It is illustrated that with this prior less error and higher frequentist coverage are obtained. The real dataset illustrates the value of the results emanating from the speculative

research.

2. The hypergeometric gamma distribution

In this section the prior class will be briefly introduced with some properties. This rich model is the univariate counterpart of the generalized matrix variate Wishart distribution of [8].

Definition 1. *The hypergeometric gamma distribution (HGD) with parameters $a_1, \dots, a_p, b_1, \dots, b_q, c, \varphi$ and n has density function*

$$f(x) = \frac{\varphi^{-\frac{n}{2}}}{\Gamma(\frac{n}{2}) {}_pF_q(\frac{n}{2}, a_1, \dots, a_p; b_1, \dots, b_q; c)} \times x^{\frac{n}{2}-1} \exp(-\varphi^{-1}x) {}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; c\varphi^{-1}x) \quad (1)$$

and is denoted by $X \sim HG(a_1, \dots, a_p, b_1, \dots, b_q, c, \varphi, n)$ for $x > 0, \varphi > 0$ and the combination of parameters, $a_1, \dots, a_p, b_1, \dots, b_q, c, \varphi$ and n , are chosen such that $f(x)$ is a proper density function, with ${}_pF_q(\cdot)$ defined in [9].

Remark 1. *Note that the r^{th} moment can be derived as*

$$E[X^r] = \frac{\varphi^r \Gamma(\frac{n}{2} + r) {}_pF_q(\frac{n}{2} + r, a_1, \dots, a_p; b_1, \dots, b_q; c)}{\Gamma(\frac{n}{2}) {}_pF_q(\frac{n}{2}, a_1, \dots, a_p; b_1, \dots, b_q; c)}$$

by using Eq. 7.522(5), p. 814 of [10].

Remark 2. *Gamma distribution (see [11]): For $p = q = 0$ and $|c| < 1$ the density function (1) reduces to*

$$f(x) = \frac{\left(\frac{(1-c)}{\varphi}\right)^{\frac{n}{2}}}{\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} \exp\left(-\frac{(1-c)}{\varphi}x\right)$$

where $x > 0, \varphi > 0$, which is the density of a gamma random variable with parameters $\frac{n}{2}$ and $\frac{(1-c)}{\varphi}$, denoted by $\text{Gamma}(\frac{n}{2}, \frac{(1-c)}{\varphi})$, since ${}_1F_0(\frac{n}{2}; c) = (1-c)^{-\frac{n}{2}}$ and ${}_0F_0(c\varphi^{-1}\sigma^2) = \exp(c\varphi^{-1}\sigma^2)$.

Remark 3. *Noncentral gamma distribution (see [12]): For $p = 0$ and $q = 1$ the density function (1) reduces to*

$$f(x) = \frac{\varphi^{-\frac{n}{2}}}{\Gamma(\frac{n}{2}) {}_1F_1(\frac{n}{2}; b_1; c)} x^{\frac{n}{2}-1} \exp(-\varphi^{-1}x) {}_0F_1(b_1; c\varphi^{-1}x)$$

where $x > 0, \varphi > 0$.

Remark 4. Definition 1 can also be rewritten as an infinite mixture of gamma density functions by using the series expansion of ${}_pF_q(\cdot)$ as follows

$$f(x) = \frac{\varphi^{-\frac{n}{2}}}{\Gamma(\frac{n}{2}) {}_pF_q(\frac{n}{2}, a_1, \dots, a_p; b_1, \dots, b_q; c)} \times \sum_{k=0}^{\infty} \frac{(a_1)_k \cdots (a_p)_k}{k! (b_1)_k \cdots (b_q)_k} c^k \varphi^{-k} x^{\frac{n}{2}+k-1} \exp(-\varphi^{-1}x),$$

therefore the reference as gamma-mixture class.

Some plots of this density function (1) are given in Figure 1 for various combinations of parameter values.

The influence of the different parameters is clearly illustrated in Figure 1. The reader will note the following interesting points:

- The density function becomes less positively skewed as the values of a , c and n increase, or the value of b decreases. For large values of a , c and n the density can become asymptotically symmetric, although large values of n might not be practically feasible and $|c| < 1$.
- The value of φ can be used to influence the kurtosis.

In the forthcoming section, we develop a subjective Bayesian analysis for the normal model by employing the HGD as the prior for the variance.

3. Bayesian analysis

Consider a sample of m observations from the normal model $N(\mu, \sigma^2)$, where both parameters are unknown, with the following likelihood function

$$L(\mu, \sigma^2 | \mathbf{x}) = (2\pi\sigma^2)^{-\frac{m}{2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^m (x_i - \mu)^2\right] \quad (2)$$

Further assume an objective prior for μ , and independently the HGD (see (1)) as the subjective prior for σ^2 such that the joint prior density function is given

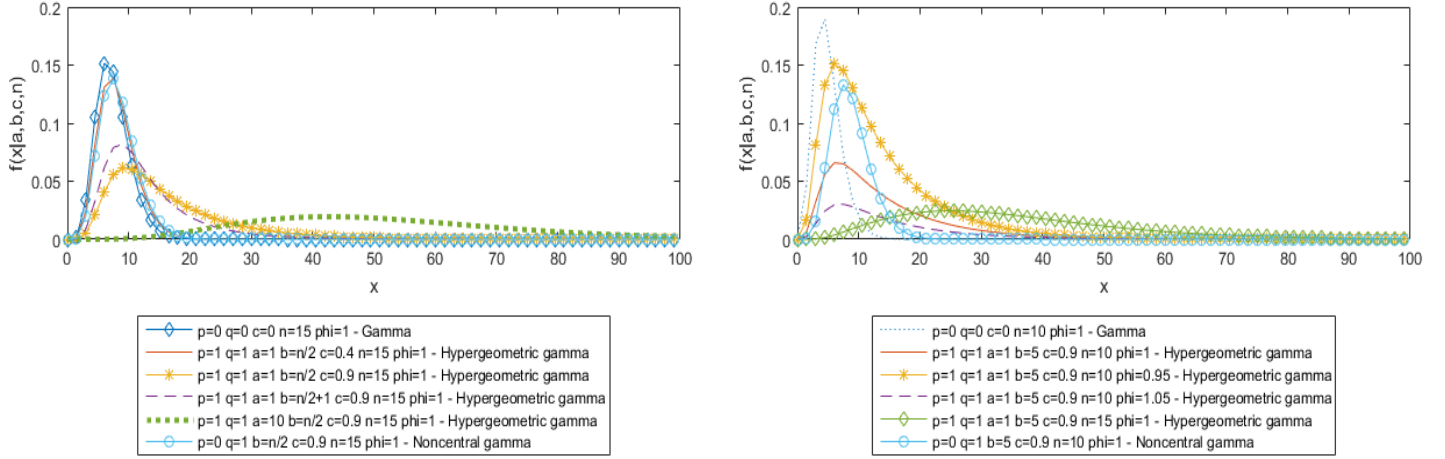


Figure 1: Density function plots for special cases of the hypergeometric gamma distribution (see (1)) - effect of a, b and c (Left), and n and φ (Right)

by

$$\pi(\mu, \sigma^2) = \frac{\varphi^{-\frac{n}{2}} (\sigma^2)^{\frac{n}{2}-1}}{\Gamma(\frac{n}{2}) {}_pF_q(\frac{n}{2}, a_1, \dots, a_p; b_1, \dots, b_q; c)} \exp(-\varphi^{-1}\sigma^2) {}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; c\varphi^{-1}\sigma^2) \quad (3)$$

where we used $\pi(\mu, \sigma^2) = \pi(\mu)\pi(\sigma^2)$ for $\pi(\mu) \equiv 1$.

Then, the joint posterior density function is obtained from (2) and (3) as follows

$$q(\mu, \sigma^2 | \mathbf{x}) \propto (\sigma^2)^{\frac{n}{2}-\frac{m}{2}-1} {}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; c\varphi^{-1}\sigma^2) \times \exp(-\varphi^{-1}\sigma^2) \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^m (x_i - \mu)^2\right] \quad (4)$$

To obtain the marginal posterior density functions, the mathematical solutions of the complicated integrals lead to unattractive expressions for the posterior density functions. We propose another novel approach to obtain the analytical expressions in situations where the usual mathematical approach leads to computationally complicated forms. To solve for the marginal posterior density

function of μ , we will obtain the following integral

$$\int_0^\infty (\sigma^2)^{\frac{n}{2} - \frac{m}{2} - 1} {}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; c\varphi^{-1}\sigma^2) \\ \times \exp(-\varphi^{-1}\sigma^2) \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^m (x_i - \mu)^2\right] d\sigma^2$$

which can be solved mathematically by using the series expansion of ${}_pF_q(\cdot)$ and the definition of the Bessel function of the third kind [10], as

$$\sum_{k=0}^{\infty} c_k K_{\frac{n}{2} - \frac{m}{2} + k} \left(2\sqrt{\sum_{i=1}^m (x_i - \mu)^2 (1-c)\varphi^{-1}} \right)$$

where $K_\nu(\cdot)$ is the Bessel function of the third kind defined by Eq.3.478(4), p.370 of [10]. This leads to a computationally challenging posterior density function which is not implementable and Gibbs sampling will have to be used to approximate this posterior density function. However, if the integrand is viewed as the product of a gamma density function and some function of σ^2 , the following theorem follows.

Theorem 1. *The marginal posterior density of μ for the normal model with prior (3) is*

$$q_1(\mu|\mathbf{x}) \\ = \frac{\Gamma\left(\frac{n}{2} - \frac{m}{2}\right)}{(2\pi m^{-1}\varphi)^{\frac{1}{2}} \Gamma\left(\frac{n}{2} - \frac{m}{2} + \frac{1}{2}\right)} \\ \times \frac{E_{\sigma_1^2} \left[\exp\left[-\frac{1}{2\sigma_1^2} \sum_{i=1}^m (x_i - \mu)^2\right] {}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; c\varphi^{-1}\sigma_1^2) \right]}{E_{\sigma_2^2} \left[\exp\left[-\frac{1}{2\sigma_2^2} (\sum_{i=1}^m x_i^2 - m\bar{x}^2)\right] {}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; c\varphi^{-1}\sigma_2^2) \right]} \quad (5)$$

where $\sigma_1^2 \sim \text{Gamma}\left(\frac{n}{2} - \frac{m}{2}, \varphi^{-1}\right)$ and $\sigma_2^2 \sim \text{Gamma}\left(\frac{n}{2} - \frac{m}{2} + \frac{1}{2}, \varphi^{-1}\right)$, provided $n > m, \varphi > 0$ and $E_{\sigma_i^2}[\cdot]$ denotes the expected value with respect to the distribution of $\sigma_i^2, i = 1, 2$.

Remark 5. *This result (5) can easily be implemented for computational use as is shown in Section 4. This methodology can be used for the derivation of*

any posterior density function if a component of the kernel of the prior is some known density function from which simulation of random variates is possible. In general, let $h(\boldsymbol{\theta}|\mathbf{x})$ be the likelihood function of $\boldsymbol{\theta} = (\theta_1, \theta_2), \theta_i \in \Omega_i, i = 1, 2$ for given \mathbf{x} . The prior density assumed is $\pi(\boldsymbol{\theta})$. Consider the marginal posterior density function of θ_i :

$$\begin{aligned} q(\theta_i, \mathbf{x}) &= \int_{\Omega_j} h(\boldsymbol{\theta}|\mathbf{x})\pi(\boldsymbol{\theta})d\theta_j \\ &= \int_{\Omega^*} g(\theta_i, \phi|\mathbf{x})f(\phi)d\phi \\ &= E_\phi(g(\theta_i, \phi|\mathbf{x})) \end{aligned}$$

with ϕ a function of θ_j , $\phi \in \Omega^*$ and $f(\phi)$ a known density function from which random variates can be simulated. This result can then be computationally calculated using the central limit theorem as

$$E_\phi(\widehat{g(\theta_i, \phi|\mathbf{x})}) = \frac{1}{n_1} \sum_{k=1}^{n_1} g(\theta_i, \phi_k|\mathbf{x})$$

where $\phi_k, k = 1, \dots, n_1$ is a random variate from the distribution with density function $f(\phi)$.

Theorem 2. The marginal posterior density of σ^2 for the normal model with prior (3) is

$$\begin{aligned} & q_2(\sigma^2|\mathbf{x}) \\ &= \left(E_{\sigma_2^2} \left[{}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; c\varphi^{-1}\sigma_2^2) \exp \left[-\frac{1}{2\sigma_2^2} \left(\sum_{i=1}^m x_i^2 - m\bar{x}^2 \right) \right] \right] \right)^{-1} \\ & \times \frac{1}{\Gamma\left(\frac{n}{2} - \frac{m}{2} + \frac{1}{2}\right) \varphi^{\frac{n}{2} - \frac{m}{2} + \frac{1}{2}}} (\sigma^2)^{\frac{n}{2} - \frac{m}{2} + \frac{1}{2} - 1} {}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; c\varphi^{-1}\sigma^2) \\ & \times \exp(-\varphi^{-1}\sigma^2) \exp \left[-\frac{1}{2\sigma^2} \left(\sum_{i=1}^m x_i^2 - m\bar{x}^2 \right) \right] \end{aligned} \quad (6)$$

with $\sigma_2^2 \sim \text{Gamma}\left(\frac{n}{2} - \frac{m}{2} + \frac{1}{2}, \varphi^{-1}\right)$ provided $n > m, \varphi > 0$.

The proofs of the aforementioned theorems can be found in Appendix A.

3.1. Bayes estimators

Under the squared error loss function the Bayes estimator is the posterior mean value. In what follows, we show that the Bayes estimator of μ is $\hat{\mu} = E[\mu|\mathbf{x}] = \bar{x}$.

The expected value of $\mu - \bar{x}$ is given by

$$\begin{aligned} E[\mu - \bar{x}|\mathbf{x}] &= \int_{-\infty}^{\infty} (\mu - \bar{x}) q_1(\mu|\mathbf{x}) d\mu \\ &= \frac{\varphi^{-\frac{1}{2}} (2\pi m^{-1})^{-\frac{1}{2}} \Gamma\left(\frac{n}{2} - \frac{m}{2}\right) \left(\Gamma\left(\frac{n}{2} - \frac{m}{2} + \frac{1}{2}\right)\right)^{-1}}{E_{\sigma_2^2} \left[\exp \left[-\frac{1}{2\sigma_2^2} \left(\sum_{i=1}^m x_i^2 - m\bar{x}^2 \right) \right] {}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; c\varphi^{-1}\sigma_2^2) \right]} \\ &\quad \times \int_{-\infty}^{\infty} (\mu - \bar{x}) E_{\sigma_1^2} \left[\exp \left[-\frac{1}{2\sigma_1^2} \sum_{i=1}^m (x_i - \mu)^2 \right] {}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; c\varphi^{-1}\sigma_1^2) \right] d\mu \end{aligned}$$

from Theorem 1. It is quite clear that the integrand is an odd function and therefore

$$\int_{-\infty}^{\infty} (\mu - \bar{x}) E_{\sigma_1^2} \left[\exp \left[-\frac{1}{2\sigma_1^2} \sum_{i=1}^m (x_i - \mu)^2 \right] {}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; c\varphi^{-1}\sigma_1^2) \right] d\mu = 0$$

We can thus conclude that $E[\mu - \bar{x}|\mathbf{x}] = 0$ and hence

$$E[\mu|\mathbf{x}] = \hat{\mu} = \bar{x} \quad (7)$$

The Bayes estimator of σ^2 from Theorem 2, is given by

$$\begin{aligned} &\hat{\sigma}^2 \\ &= E[\sigma^2|\mathbf{x}] \\ &= \left(E_{\sigma_2^2} \left[{}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; c\varphi^{-1}\sigma_2^2) \exp \left[-\frac{1}{2\sigma_2^2} \left(\sum_{i=1}^m x_i^2 - m\bar{x}^2 \right) \right] \right] \right)^{-1} \\ &\quad \times \int_0^{\infty} \frac{1}{\Gamma\left(\frac{n}{2} - \frac{m}{2} + \frac{1}{2}\right) \varphi^{\frac{n}{2} - \frac{m}{2} + \frac{1}{2}}} (\sigma^2)^{\frac{n}{2} - \frac{m}{2} + \frac{3}{2} - 1} {}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; c\varphi^{-1}\sigma^2) \\ &\quad \times \exp(-\varphi^{-1}\sigma^2) \exp \left[-\frac{1}{2\sigma^2} \left(\sum_{i=1}^m x_i^2 - m\bar{x}^2 \right) \right] d\sigma^2 \\ &= \frac{\left(\frac{n}{2} - \frac{m}{2} + \frac{1}{2}\right) \varphi E_{\sigma_3^2} \left[{}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; c\varphi^{-1}\sigma_3^2) \exp \left[-\frac{1}{2\sigma_3^2} \left(\sum_{i=1}^m x_i^2 - m\bar{x}^2 \right) \right] \right]}{E_{\sigma_2^2} \left[{}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; c\varphi^{-1}\sigma_2^2) \exp \left[-\frac{1}{2\sigma_2^2} \left(\sum_{i=1}^m x_i^2 - m\bar{x}^2 \right) \right] \right]} \end{aligned} \quad (8)$$

with $\sigma_2^2 \sim \text{Gamma}\left(\frac{n}{2} - \frac{m}{2} + \frac{1}{2}, \varphi^{-1}\right)$ and $\sigma_3^2 \sim \text{Gamma}\left(\frac{n}{2} - \frac{m}{2} + \frac{3}{2}, \varphi^{-1}\right)$.

3.2. Special cases

Referring to Remark 2 the posterior distributions for the gamma prior can then be obtained directly from (5) and (6) as

$$q_1(\mu|\mathbf{x}) = \frac{(1-c)^{\frac{1}{4}} 2^{\frac{1}{4}} \left(\sum_{i=1}^m (x_i - \mu)^2\right)^{\frac{n}{4} - \frac{m}{4}} K_{\frac{n}{2} - \frac{m}{2}} \left(2\sqrt{\sum_{i=1}^m (x_i - \mu)^2 (1-c)\varphi^{-1}}\right)}{(2\pi m^{-1})^{\frac{1}{2}} \left(\sum_{i=1}^m x_i^2 - m\bar{x}^2\right)^{\frac{n}{4} - \frac{m}{4} + \frac{1}{4}} K_{\frac{n}{2} - \frac{m}{2} + \frac{1}{2}} \left(2\sqrt{\left(\sum_{i=1}^m x_i^2 - m\bar{x}^2\right) (1-c)\varphi^{-1}}\right)}$$

and

$$\begin{aligned} q_2(\sigma^2|\mathbf{x}) &= \frac{\left(\frac{-\sum_{i=1}^m x_i^2 - m\bar{x}^2}{2c\varphi^{-1}}\right)^{-\frac{n}{4} + \frac{m}{4} - \frac{1}{4}}}{2K_{\frac{n}{4} - \frac{m}{4} + \frac{1}{4}} \left(2\sqrt{-\left(\sum_{i=1}^m x_i^2 - m\bar{x}^2\right) c\varphi^{-1}}\right)} (\sigma^2)^{\frac{n}{2} - \frac{m}{2} + \frac{1}{2} - 1} \\ &\quad \times \exp\left(-\frac{1}{2\sigma^2} \left(\sum_{i=1}^m x_i^2 - m\bar{x}^2\right)\right) \exp\left[-(1-c)\varphi^{-1}\sigma^2\right] \end{aligned}$$

similar to the results of [7] where $K_\nu(\cdot)$ is the Bessel function of the third kind defined by Eq.3.478(4), p.370 of [10]. The Bayes estimators are as follows

$$\hat{\mu} = \bar{x}$$

and

$$\begin{aligned} \hat{\sigma}^2 &= \left(\frac{-\sum_{i=1}^m x_i^2 - m\bar{x}^2}{2c\varphi^{-3}}\right)^{\frac{1}{2}} \\ &\quad \times \frac{K_{\frac{n}{4} - \frac{m}{4} + \frac{3}{4}} \left(2\sqrt{-\left(\sum_{i=1}^m x_i^2 - m\bar{x}^2\right) c\varphi^{-1}}\right)}{K_{\frac{n}{4} - \frac{m}{4} + \frac{1}{4}} \left(2\sqrt{-\left(\sum_{i=1}^m x_i^2 - m\bar{x}^2\right) c\varphi^{-1}}\right)} \end{aligned}$$

Complete proofs are given in Appendix B.

4. Illustrations

4.1. Evaluation using the proposed methodology (see Remark 5)

In this section we show that the analytical posterior density functions (5) and (6), as well as the Bayes estimators (7) and (8) (derived using the proposed methodology) for the hypergeometric gamma prior with $p = 1$ and $q = 1$,

produces at least as good results as Gibbs sampling when approximated from a sample.

For this purpose, we develop a Gibbs sampling scheme with a Metropolis-Hastings algorithm to simulate posterior samples of μ and σ^2 . A sample of 1000 posterior values are simulated with a burn-in of 10%. The empirical histograms are obtained from these samples and can be observed in Figure 2.

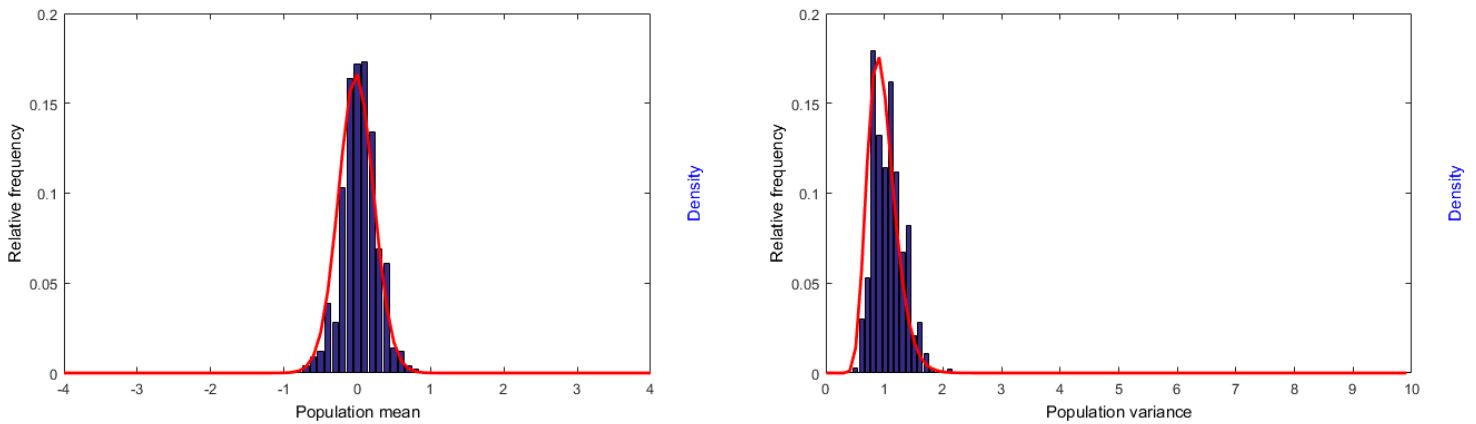


Figure 2: Analytical posterior density functions (5) (Left) and (6) (Right) and histograms based on Gibbs sampling for μ (Left) and σ^2 (Right)

Remark 6. *The computational time to construct the histograms is 812.56 seconds, whereas the curves using the proposed analytical expressions is 68.52 seconds.*

It is clear that the analytical posterior density functions in (5) and (6) are supported by results obtained from the Gibbs sampling method. The advantage of the proposed methodology is that no specialized software or advanced programming abilities are needed. This approach provides an alternative to approximations such as MCMC-MH sampling and INLA, since the implementation is computationally efficient, especially for complicated prior structures such as the case under consideration.

4.2. Bayesian performance study

The newly developed results for (1) as prior, will be applied in a simulation context and compared with other results from literature. Note that this is not an exhaustive simulation study but just an illustration of the proposed results.

4.2.1. The choice of hyperparameter values

The choice of the prior parameters is a crucial aspect of any subjective Bayesian analyses. We propose an intuitive methodology on the parameter values with the use of this specific prior structure similar to Duran and Booker [13]. Assuming the hypergeometric gamma distribution (see Definition 1 with $p = 1$ and $q = 1$) as one of the priors for σ^2 , with the prior density function

$$\pi(\sigma^2) = \frac{\varphi^{-\frac{n}{2}} (\sigma^2)^{\frac{n}{2}-1}}{\Gamma(\frac{n}{2}) {}_2F_1(\frac{n}{2}, a, b; c)} \exp(-\varphi^{-1}\sigma^2) {}_1F_1(a; b; c\varphi^{-1}\sigma^2) \quad (9)$$

where ${}_1F_1(\cdot)$ is the confluent hypergeometric function and ${}_2F_1(\cdot)$ is the Gauss hypergeometric function with $|c| < 1$ (see [9]), with first moment

$$E[\sigma^2] = \frac{\varphi\Gamma(\frac{n}{2} + 1) {}_2F_1(\frac{n}{2} + 1, a; b; c)}{\Gamma(\frac{n}{2}) {}_2F_1(\frac{n}{2}, a; b; c)}$$

Consider the prior is constrained under the following belief

$$H_0 : \sigma^2 = \sigma_0^2$$

Setting the expected value of σ^2 equal to σ_0^2 yields

$$\begin{aligned} \sigma_0^2 &= \frac{\varphi\Gamma(\frac{n}{2} + 1) {}_2F_1(\frac{n}{2} + 1, a; b; c)}{\Gamma(\frac{n}{2}) {}_2F_1(\frac{n}{2}, a; b; c)} \\ &= \frac{(\frac{n}{2})\varphi {}_2F_1(\frac{n}{2} + 1, a; b; c)}{{}_2F_1(\frac{n}{2}, a; b; c)} \end{aligned} \quad (10)$$

The ratio $\frac{{}_2F_1(\frac{n}{2}+1, a; b; c)}{{}_2F_1(\frac{n}{2}, a; b; c)}$ is a slow varying function as can be seen in Figure 3.

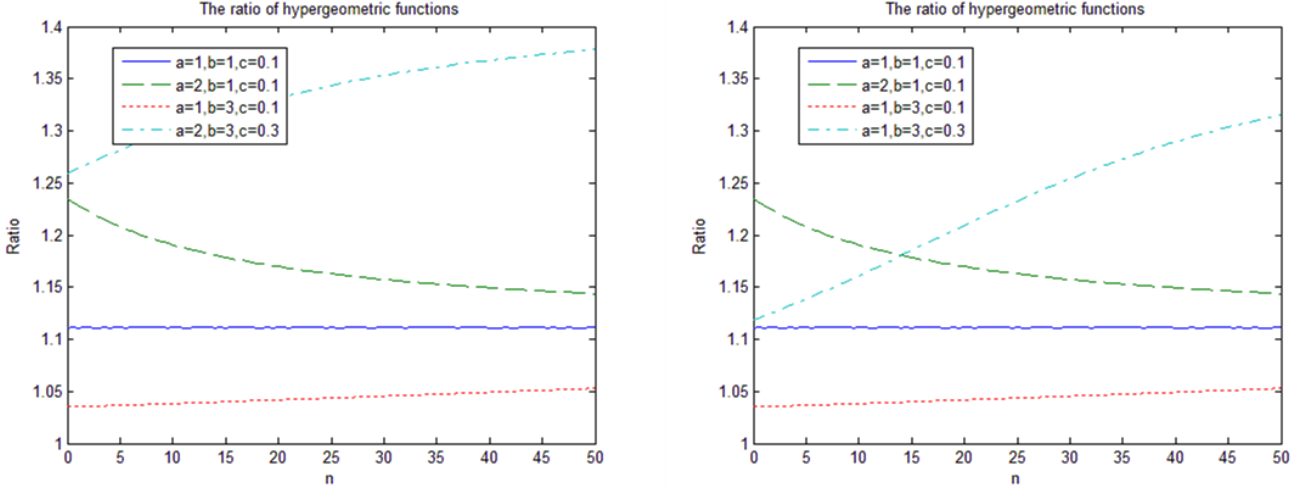


Figure 3: The ratio, $\frac{{}_2F_1(\frac{n}{2}+1, a; b; c)}{{}_2F_1(\frac{n}{2}, a; b; c)}$, for various parameter values

Hence

$$\sigma_0^2 \approx \left(\frac{n}{2}\right)\varphi$$

with $n > m$, where m is the sample size. This enables us to choose φ in an informed manner, based on the choice of n .

4.2.2. Simulation study

For our purposes, a normal sample of size 18 was simulated with parameters $\mu = 0$ and $\sigma^2 = 1$. The hyperparameters are chosen according to the above methodology as $n = 19$, $\varphi = 0.105$, $a = 1$, $b = 2$ and $c = 0.01$ such that $E(\sigma^2) = 0.9$. The choice of $\sigma_0^2 = 0.9$ not being exactly $\sigma^2 = 1$ illustrates that this prior performs well even under an incorrect prior belief. The sample estimates which are the MLE's are given by

$$\widehat{\mu} = \bar{x} = -0.0166, \quad \widehat{\sigma^2} = s^2 = 0.8397$$

The Bayes estimates calculated using the sample, from (7) and (8), respectively, are

$$\widehat{\mu} = \bar{x} = -0.0166, \quad \widehat{\sigma^2} = s^2 = 0.9701$$

It is evident from the comparison of the estimates that the Bayes estimate for σ^2 is closer to the target parameter value than the MLE. The following figure displays the analytical posterior density functions as in (5) and (6) with $p = 1$ and $q = 1$

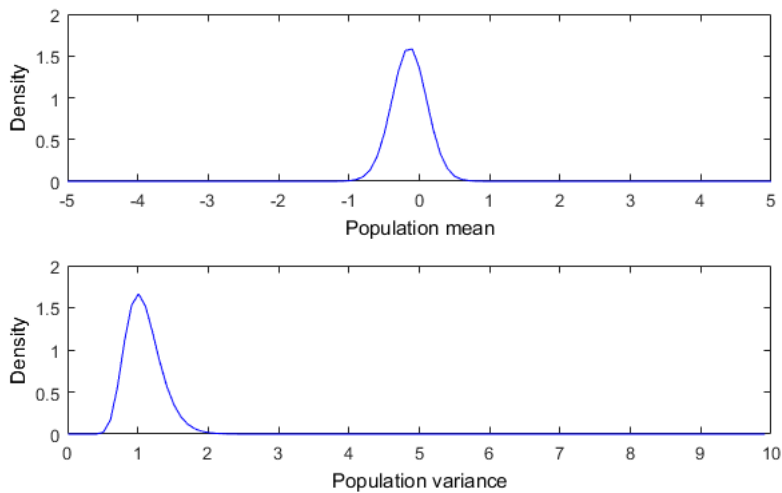


Figure 4: Analytical posterior density functions of μ and σ^2

4.2.3. Comparative simulation study

In this section a normal sample of size m is simulated with known mean and variance and the priors under consideration are the well-known inverse-gamma prior (see [11]), as well as the gamma prior (see Remark 1) and the hypergeometric gamma prior (see Definition 1 with $p = 1$ and $q = 1$) as special cases of this gamma class. The newly proposed estimators ((7) and (8) with $p = 1$ and $q = 1$) for μ and σ^2 , respectively, are calculated and compared with the two other Bayes estimators as well as the MLE's in terms of coverage probabilities and median credible interval width.

Remark 7. *The hyperparameters of the inverse gamma and gamma priors were chosen in an analogous way based on the prior belief*

$$H_0 : \sigma^2 = \sigma_0^2$$

For our purposes $\sigma_0^2 = 0.9$. Four combinations of hyperparameters are investigated and summarized in Table 2.

Note that combination 4 in Table 2 results in a hypergeometric gamma prior with an approximate prior expected value of 4.75, while the prior belief is 0.9 for the other combinations.

Combination	Inverse-Gamma prior	Gamma prior	Hypergeometric gamma prior
1	0.87(0.9)	0.91(1)	0.99(0.8)
2	0.88(0.9)	0.92(1)	0.97(0.8)
3	0.90(0.9)	0.94(1)	0.98(0.8)
4	0.88(0.9)	0.92(1)	0.89(1.3)

Table 1: Coverage probabilities (median credible interval width) for different combinations

The coverage probabilities calculated under the hypergeometric gamma prior is the highest with a low median width of the credible interval for combinations 1 to 3, thus the Bayes estimators derived under this structure are good. It is interesting that even in combination 4, the hypergeometric gamma prior still performs satisfactorily well. The bias and mean squared error for each prior and each sample were calculated, for combinations 1 and 4, and are depicted in Figure 5.

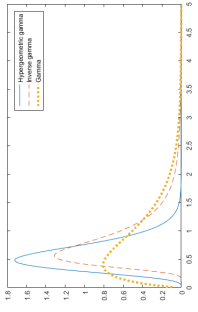
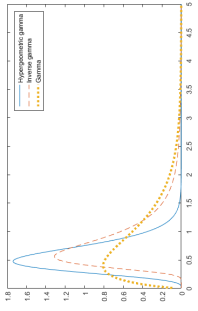
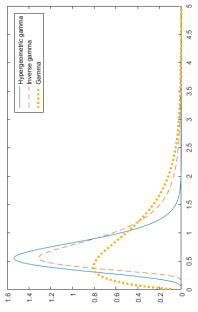
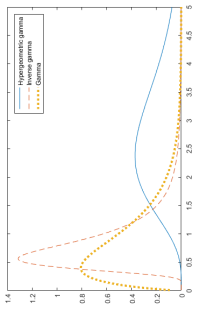
Combination	1	2	3	4
Ⓐ	$\alpha_1 = 4.33, \beta_1 = 3$	$\alpha_1 = 4.33, \beta_1 = 3$	$\alpha_1 = 4.33, \beta_1 = 3$	$\alpha_1 = 4.33, \beta_1 = 3$
Ⓑ	$\alpha_2 = 1.8, \beta_2 = 2$	$\alpha_2 = 1.8, \beta_2 = 2$	$\alpha_2 = 1.8, \beta_2 = 2$	$\alpha_2 = 1.8, \beta_2 = 2$
Ⓒ	$\varphi = 0.105, c = 0.02, n = 19, a = 1, b = 1$	$\varphi = 0.105, c = 0.02, n = 19, a = 1, b = 2$	$\varphi = 0.105, c = 0.02, n = 19, a = 10, b = 1$	$\varphi = 0.5, c = 0.1, n = 19, a = 1, b = 2$
				

Table 2: Influence of hyperparameters on the prior density functions; Ⓐ - Inverse gamma prior (- -), Ⓑ - Gamma prior (...), Ⓒ - Hypergeometric gamma prior (-)

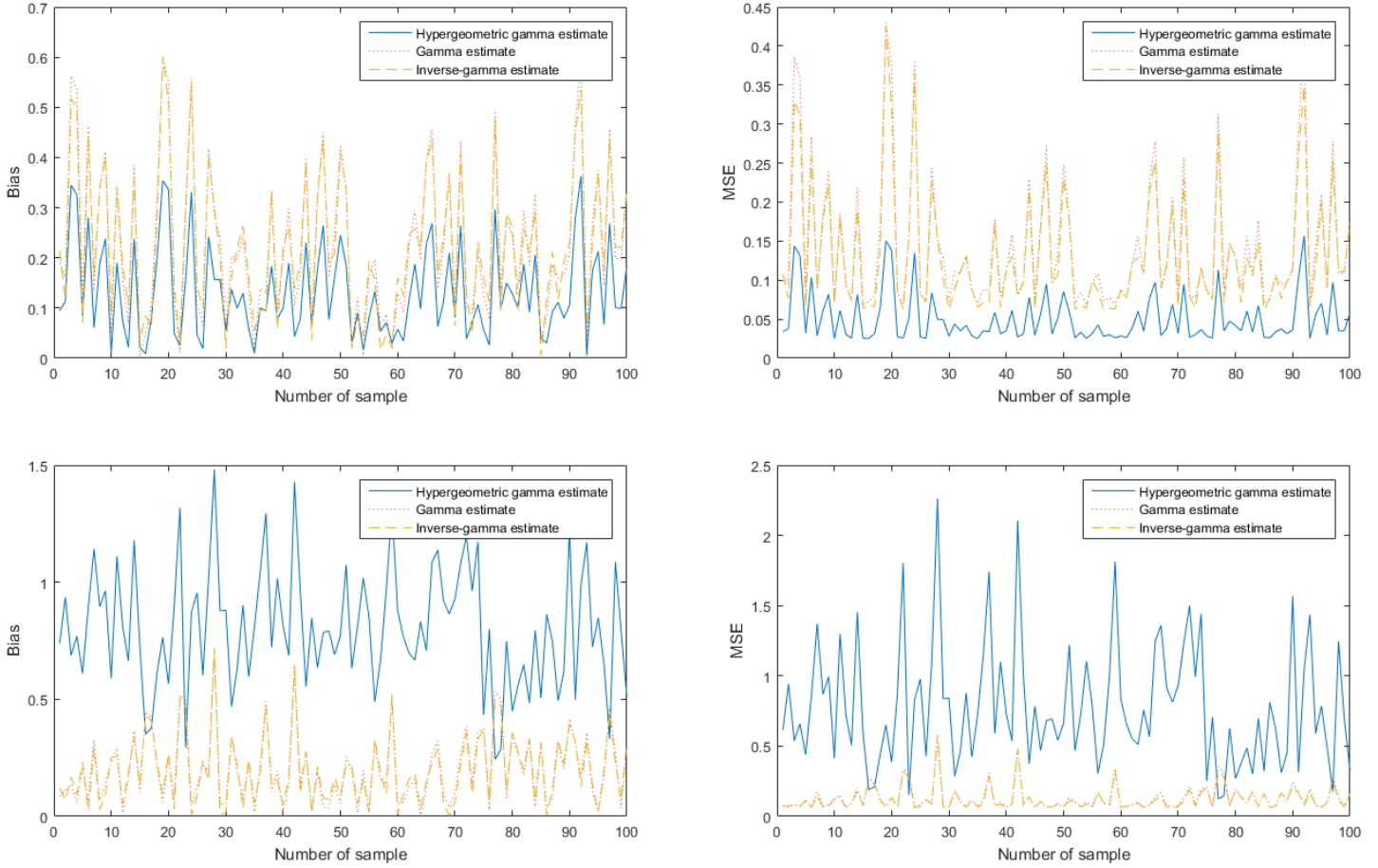


Figure 5: Bias (Left) and Mean squared error (MSE) (Right) for 100 samples of combination 1 (top row) and combination 4 (bottom row) (Hypergeometric gamma prior (-), gamma prior (...) and inverse-gamma prior (- -))

The good performance of the hypergeometric gamma prior for combination 1 is evident from Figure 5 since the bias and MSE are smaller for this prior than for the inverse-gamma prior. More precise, and hence more accurate estimates are obtained for combination 1, compared to combination 4. This is supported by Table 1. From Table 1 and Figure 5, the results of [3] and [7] are again apparent regarding the better performance of the gamma prior when compared

to the inverse-gamma prior.

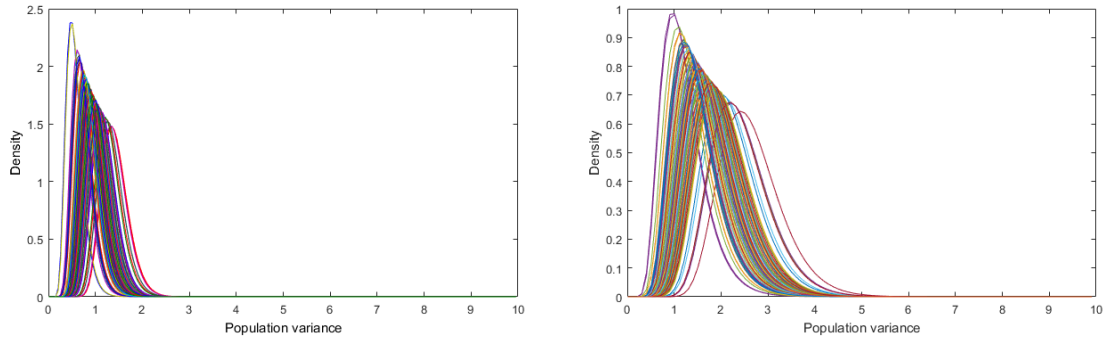


Figure 6: Analytical posterior density function of σ^2 for 100 samples using combination 1 (Right) and combination 4 (Left)

The marginal posterior density function (6) for the hypergeometric gamma prior, with $p = 1$ and $q = 1$, of σ^2 for the 100 different samples can be viewed in Figure 6. It can be remarked that this density function (6) is quite robust with respect to sampling in the sense that the shape and scale is not heavily influenced by a change in the sample, even for a small sample as in this paper. It is also evident that the posterior is not dominated by the shape of the prior by comparing Table 2 and Figure 6.

4.3. Real dataset

A forester wishes to estimate the volume of merchantable timber in a population of trees, based on a sample [14, p.28]. To this purpose he selected 31 trees and measured their volume. Since this is a small sample, most probably not random due to location constraints, we propose a Bayesian approach with a hypergeometric gamma prior for the estimation of the mean and variance. Inverse gamma and gamma priors are also considered. The data is found to be non-normal and a log-transformation corrected this. Let X be the volume of the tree and $Y = \log(X)$ be the log-transformed variable, then $Y \sim N(\mu, \sigma^2)$. The prior information is specified as $E[\sigma^2] = 0.3$ based on expert information.

Five different priors are assumed for σ^2 as summarized in Table 3 and illustrated in Figure 7. The hyperparameters of the hypergeometric gamma priors are chosen according to the methodology in Section 4.2.1. The likelihood, priors

Prior	Hyperparameters	Bayes estimate	95% Credible interval	Width
Hypergeometric gamma 1 *	$\varphi = 0.01875, n = 32$	0.2933	(0.21;0.385)	0.175
Hypergeometric gamma 2 *	$\varphi = 0.1, n = 6$	0.2945	(0.195;0.425)	0.23
Hypergeometric gamma 3 *	$\varphi = 0.5, n = 1.2$	0.2976	(0.195;0.445)	0.25
Inverse gamma	$\alpha_1 = 3.5, \beta_1 = 11.5$	0.3107	(0.195;0.465)	0.26
Gamma	$\alpha_2 = 3.3, \beta_2 = 10$	0.3125	(0.19;0.41)	0.22

Table 3: Hyperparameters and posterior characteristics under the considered priors

* $a=1, b=1, c=0.01$

and posterior density functions of σ^2 for this dataset is displayed in Figure 7.

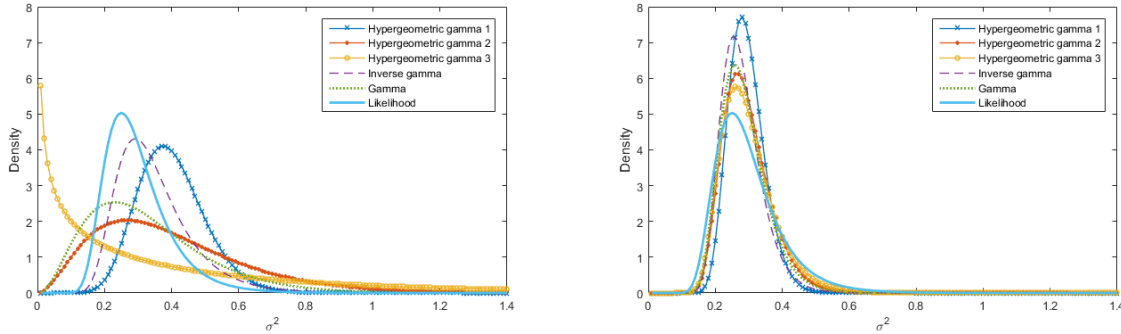


Figure 7: Prior density functions (Left) and analytical posterior density functions (Right) of σ^2

It was illustrated in this section that different prior structures still lead to a more concentrated posterior which in turn shows that the prior is not in conflict with the likelihood (see [15]), and more than one prior may be appropriate. McElreath [16] stated: "People commonly ask what the correct prior is for a given analysis. The question sometimes implies that for a given set of data,

there is a uniquely correct prior that must be used, or else the analysis will be invalid. This is a mistake. There is no more a uniquely correct prior that there is a uniquely correct likelihood.”

The advantage of the hypergeometric gamma prior is clearly seen in Figure 7 for the real dataset. The resulting credible intervals are narrower even though the priors are vaguer than the inverse gamma prior, leading to more precise estimates. The gamma prior still performs better than the inverse gamma prior.

5. Conclusion

In this paper we proposed a new class, where this hypergeometric gamma distribution is a prior for the the univariate normal model. The analytical expressions of the posterior density functions and Bayes estimators were simplified by identifying complex expressions as expected values. This methodology was generalized in Remark 5. This approach simplifies computations when informative priors are used with rich structure and can pave the way for the use of this in several statistical analyses, amongst others regression analysis. This new gamma model delivered higher accuracy and higher coverage than the other priors. The advantages of using the analytical expressions instead of MCMC sampling were illustrated in terms of computational time. Additionally, no knowledge of MCMC sampling and Metropolis-Hastings algorithms are needed for the application of these results. In the simulation studies and the analysis of the real dataset it was shown that this new gamma model optimizes the modeling of real phenomena.

6. Acknowledgements

The authors would like to hereby acknowledge the support of the StatDisT group. This work is based upon research supported by the National Research foundation, Grant (Re:CPRR13090132066 No 91497) and the vulnerable discipline-

academic statistics (STAT) fund. We would like to thank the reviewers and the associate editor for the valuable comments and recommendations.

References

- [1] H. Raïffa, R. Schlaifer, Applied statistical decision theory, Studies in managerial economics, Division of Research, Graduate School of Business Administration, Harvard University, 1961. URL: <https://books.google.co.za/books?id=wPBLAAAAMAAJ>.
- [2] S. Bhattacharya, A. Saxena, Multivariate modified Bessel distribution, *Estadistica* 41 (1989).
- [3] J. E. Griffin, P. J. Brown, Inference with normal-gamma prior distributions in regression problems, *Bayesian Anal.* 5 (2010) 171–188.
- [4] A. Gelman, et al., Prior distributions for variance parameters in hierarchical models, *Bayesian analysis* 1 (2006) 515–534.
- [5] A. Gelman, Bayes, Jeffreys, prior distributions and the philosophy of statistics, *Statistical Science* 24 (2009) 176–178.
- [6] N. Sebastian, A generalized gamma model associated with a Bessel function, *Integral transforms and Special functions* 22 (2011) 631–645.
- [7] J. Van Niekerk, A. Bekker, A. Arashi, J. J. J. Roux, Subjective Bayesian analysis of the elliptical model, *Communications in Statistics - Theory and Methods* 44 (2015) 3738–3753.
- [8] J. J. J. Roux, A characterization of multivariate distributions, *South African Statistical Journal* 5 (1971) 27–36.
- [9] A. Erdelyi, W. Magnus, F. Oberhettinger, F. G. Tricomi, Higher transcendental functions I, McGraw-Hill Inc., United States of America, 1953.
- [10] I. S. Gradshteyn, I. M. Ryzhik, Table of integrals, series and products, Academic Press, United States of America, 2007.

- [11] M. DeGroot, Optimal statistical decisions, McGraw-Hill Inc., United States of America, 1970.
- [12] N. Johnson, S. Kotz, N. Balakrishnan, Continuous univariate distributions, number v. 2 in Wiley series in probability and mathematical statistics: Applied probability and statistics, Wiley & Sons, 1995. URL: <https://books.google.co.za/books?id=0QzvAAAAMAAJ>.
- [13] B. Duran, J. Booker, A Bayes sensitivity analysis when using the beta distribution as a prior, IEEE transactions on reliability 37 (1988) 239–247.
- [14] A. Grafen, R. Hails, R. Hails, R. Hails, Modern statistics for the life sciences, volume 123, Oxford University Press Oxford, 2002.
- [15] E. Lesaffre, A. B. Lawson, Bayesian biostatistics, John Wiley & Sons, 2012.
- [16] R. McElreath, Statistical Rethinking: A Bayesian course with examples in R and Stan, volume 122, CRC Press, 2016.